

Chapter 1

Introduction

By the turn of the century, the analysis and management of risk had escaped the bounds of professional concern and scholarly expertise to claim a prominent place in the public consciousness. The Information Society had become the Risk Society, and the arcane wizardry of actuaries and statisticians became common, almost essential features of popular mass media fare. Estimates of probability based on analyses of events in the past have come to dominate decisions about the paths we should take in the future. Nothing worthy of our attention can avoid an assessment of chance.

There are undoubtedly a great many benefits that have been derived from the application of probability and statistics to an unimaginable variety of problems and concerns. In the view of some, they have “become the very guide of life,” and they have done so in extremely powerful ways. Indeed, in their almost routine application, they “have called new objects into being, coined new values, and established new standards of rationality and new claims to authority” (Gigerenzer et al. 1989: 270). Not all of these outcomes are the objects of praise, however. Indeed, this book is about the ways in which some uses of probability and statistics serve to reproduce and reinforce disparities in the quality of life that different sorts of people can hope to enjoy.

For some people, it seems that if it weren't for bad luck, they wouldn't have any luck at all. The impact of bad luck seems to cumulate rapidly over time, such that a bit of bad luck early in life increases the probability that losses will mount, and the gambler's dream of breaking even, or getting ahead in the game eventually gives way to despair.

We are just beginning to understand how much initial positions play in the ways our lives develop. How well we do in the natural lottery that distributes genetic endowments at birth helps to determine how race, gender, and social class combine in unimaginably chaotic ways to move us down different paths along the unmapped roads of life. We are just beginning to appreciate the complexity in the mutually reinforcing and limiting systems that generates an array of opportunity structures or life chances that ultimately determine who, how, and what we are when we reach the end of our journey.

This book examines the ways in which public policy and private action further shapes the design of a variety of games that we may or may not see as fair. Policy decisions, informed by statistical analysis, shape the opportunities people face in the markets for education, employment and health. Strikingly similar analyses also determine what we can expect when we take our chances with the courts and the criminal justice system. Special attention is paid to the role played by an

actuarial logic that informs routine decisions about access to financial resources, including insurance, that govern far more than access to credit.

The mass media play a critically important role in shaping the ways in which we understand the role of chance and luck in our lives, and in the lives of others. The ways in which the media frame the life chances that different groups confront helps to determine whether the public supports, opposes, or ignores proposals to modify public and private activity in these critical areas. This book provides an analysis of tendencies within the media that ironically serve to reinforce hardship and suffering in the lives of people that many investigative journalists had intended to lend a helping hand.

Challenging the actuarial logic that shapes the distribution of life chances in society is an exceptionally difficult challenge. This book represents the first draft of a declaration of independence from its imperialistic grasp, and a call to oppose its spread.

What's Luck Got To Do with It?

Unpredictability is a fundamental characteristic of what we commonly understand as luck. Luck is said to thrive “in the gap between probability and actuality, between what can reasonably be expected ... and what actually occurs” (Rescher 1995: 35). We understand that when luck, or chance is really in charge, prediction is bound to fail. We might think, then, that any success in our efforts to predict the future should reduce the role of chance. At the same time, we also recognize that knowledge of the future really has to be distinguished from the actions we take to bring it into being.

Although when Karl Marx (1983) referred to men making their own history, but not making it as they pleased, he was commenting primarily on the constraints on our choices that have been established by choices made by others in the past. Our ability to make our futures is also constrained by the actions that others have not yet made, but will make in the future. And, while Karl Marx and Anthony Giddens both emphasize the fact that our futures are shaped by the actions of goal seeking individuals, Giddens (1984: 5) was more aware of the ways in which constraints on “knowledgeability” puts knowledge of the future beyond our reach.

We are not only limited in our ability to foresee the unanticipated consequences of our actions, but we are limited in our ability to make well informed choices because of the limits on our knowledge of the conditions under which we act. Just as we are not fully aware of the motivations that guide our own choices, we are largely ignorant of the motivated choices that others are, or will be making. Some of those choices are bound to change the circumstances shaping the opportunities and constraints that our own actions were meant to control.

This book is about the use of predictive technologies to shape the futures that people face in ways that no longer invite comments about misfortune or bad luck. Although we may not talk about the role of chance in many of the

situations we observe, we are likely to raise questions about the role that justice and considerations of fairness have played with regard to the distribution of the outcomes that result from the choices that powerful actors have made.

Truly random, or chance distributions tend not to be assessed in terms of fairness. Games of chance are enjoyed in part because of the pleasures derived from a bit of unexpected good luck. Even when the impact of chance is tempered by skill, as in card games or even in the game of golf, the unexpected arrival of good fortune still tends to be enjoyed. However, when the deck is stacked, or marked, or the dice are craftily weighted, the fun quickly leaves the room. We feel as though we have been cheated of the benefits that were supposed to come on the winds of chance.

The same is true in the game of life. If the games in which we must compete are in some sense fixed, such that the outcomes tend to favor the house, or some other group of players, moral outrage becomes the order of the day. Perhaps, this is what we really mean when we say "life is unfair." This is not simply an articulation of the view that we do not necessarily deserve everything that comes our way. It is also a suggestion that for some people, the odds may have been stacked against them, not merely by chance, but because strategic actors, informed by statisticians have changed the rules of the game.

It is also true that what people believe about the role of luck, fate, or chance also plays a central role in how they make their way through life. A fundamental distinction between people in terms of their beliefs regarding the "locus of control" over their lives (Rotter 1966) has been shown to be a powerful influence over the choices that they make, including how they respond to both opportunity and constraint. Although the underlying concept has been extended considerably, the primary distinction between those who believe that external factors determine their well being, and those who believe that they are the masters of their fate has been put to use in a broad range of behavioral analyses.

Even though relatively objective statistics may suggest that differences in exposure to hazardous risk are relatively slight, people in lower status positions tend to be more fearful of the myriad hazards in their environment. Some recent work suggests that the influence of status is modified to a great extent by differences in worldview, such that those with the perspectives most common among white males are also the least concerned about risks in their environment (Flynn et al. 2006). Those least concerned about risks also tend to believe that victims get what they deserve.

Consider the ways in which luck, chance, fate or fortune help to shape the quality of life that each of us may enjoy. Many believe that who we are, or who we become is both known and predetermined by the author of some grand design. However, because we don't have reliable knowledge of that plan, most of us would grant that the kind of persons we will become is unpredictable, and to some degree, subject to the influence of luck or chance. Certainly our natural talents and abilities can be understood as the product of a "natural lottery." And, although we may get to play an active role in shaping the kinds of people we will become,

a great many of the challenges and opportunities we face along the way are also shaped to a considerable extent, by the hand of fate. Of course, it's true that bad luck in the distribution of natural endowments can be overcome by good fortune in the environments that surround us. This fact helps to reinforce the impression that luck will always play some role in shaping our individual and collective futures.

One way of thinking about the possible futures that are shaped by luck as well as by the actions of others is in terms of "life chances." Life chances are understood in terms of the options and the "ligatures" or constraints that are inherent in the structures of relations and associations that define our position within society (Dahrendorf 1979). That is, if we are to understand the ways in which the use of predictive models shape the futures of the seemingly unfortunate, we have to learn quite a bit more about the individual and institutional routines that generate cumulative disadvantages for these people one decision at a time.

And, while it is important to provide some clarity with regard to the consequences that flow from individual decisions to deny or to exclude, it is even more critical to understand the strategies, incentives, and rationales of those who make those decisions in the first place.

Coming to Terms with Chance

Ian Hacking (1990: 4) suggests that probability is "*the philosophical success story of the first half of the twentieth century.*" He makes this claim in part on the basis of yet another claim that few will challenge: that most public decisions these days are made on the basis of some analysis of data that will be discussed in terms of probability and chance.

Among the many contributions that Hacking makes to our thinking about the role that probability and statistics have come to play in our lives is his call for paying special attention to the ways in which the creation of categories, including categories of people, also works to shape the lives of those who thereby come to be measured. This fundamental insight has been extended with regard to the classification of people by race, class, and social position, as well as by the jobs they hold and diseases that alter the quality of their lives (Bowker and Star 1999).

Hacking and others have offered critical histories of the place of numbers in our lives. They remind us of the role that the history of ideas about what are the normal, or central tendencies within distributions of things, often invites even more consideration of departures or deviations from those norms. How we evaluate people, places and things in terms of their departures from what we have defined as the norm, is often a fundamental determinant of the positions they will come to occupy in still other distributions that we have yet to consider.

Although Hacking did take note of the ways in which concerns about deviations from the norm within population groups led to the development of a eugenics movement, he actually minimized the role that Karl Pearson and Frances Galton played within it (Hacking 1990: 120–21). Fortunately, another, more focused

history of social statistics (Zuberi 2001), examines the links between a desire to “improve the human race” by selective breeding and more contemporary debates about the genetic basis of gaps in measured intelligence between whites and African Americans (Rushton and Jensen 2005).

Part of the contribution to statistical thinking that Pearson, Galton and others have made to our current concerns is reflected in the ways we have come to think about difference. Routinely we ask whether some measured or estimated difference is “real” or a merely a difference due to chance. This is the question that confronts researchers engaged in experiments as well as those seeking to capture the vagaries of public opinion. Unfortunately, our understanding of what these “statistically significant” differences really mean or imply is not as complete or as well distributed as many of us think it should be.

This failure to understand the meaning of difference is especially troublesome, as we have come to rely on an assumption of understanding on the part of those whose responsibility it is to develop and implement important public policies. A failure to understand the meaning of differences due to chance limits our ability to identify individual and institutional acts of racial discrimination that may have brought about these distinctions in the first place (National Research Council 2004).

Misapprehension of the true meaning of “except by chance” also includes a widespread failure to fully appreciate the importance of those things that we believe to be extremely unlikely. Nassim Taleb is a statistical thinker who occupies an important position on the margins of those who are primarily engaged with the normal and its occasional, but predictable departures. He argues for greater consideration of consequences likely to flow from the occurrence of the extreme statistical outliers that he refers to as “black swans” because they were assumed not to exist before they suddenly appeared (Taleb 2007).

We think of statistical black swans when we reflect on the impact of unanticipated events, like the attack on the twin towers of the World Trade Center in 2001. While such events can be accommodated within explanatory models after the fact, they are simply unimaginable, and therefore absent from the models that were in widespread use before they occurred. The impact of such events can be quite substantial, in part because of the ways in which responses to the event may themselves set other poorly considered processes into motion. These include the misguided and socially disruptive attempts to identify terrorists and then predict their attacks on U.S. targets (Schneier 2003).

The economic crisis that began to emerge in 2007 was also not expected, although recessions are a familiar cyclical feature of capitalist business cycles (Sherman 1987). The widespread suffering that ensued was not the product of a well coordinated terrorist plot, but was instead the result of an economic bubble that let the air out of the global economy. Neither crisis was predicted. And, in both cases, a government’s defenses were tested and found wanting, while its advisors barely knew where to begin.

Prediction versus Explanation

Although the statistical methods used to demonstrate the existence of racial and economic disparity, and perhaps, even to identify the relative contribution that different factors might play in the underlying process are similar to those used in the analytical models that guide public policy choices, the consequences that flow from these uses are quite different. Both analyses are likely to lead to decisions that change the distribution of life chances for some individuals. However, as I will argue throughout this book, the use of these techniques as a form of predictive intelligence, or a guide to rational choice, will do far more to generate disparity than their use for the assignment of blame or responsibility will help to compensate for the sins of the past. I believe that this is true, despite the fact that we have a greater chance of being correct in our assessments of the past than we have with regard to our predictions of the future.

If we are careful about what we really mean when we use statistics in an effort to explain, discover, or demonstrate how some social or behavioral system actually works, we would have to admit that we are nearly always using data gathered in the past to evaluate predictions about what also occurred in the past. These explanations are based on retrospective assessments, rather than predictions about what will occur in the future.

Many of the so-called explanatory models rely on statistical regression or other techniques that assess the association between historical measures or data. Although analysts tend to talk about their estimates in causal terms, these models are not truly causal, but instead they are correlational. Statements about effects, or even effect size simply means that variation in some measured factor “is associated to some degree” with something else (Holland 1986). Still, the use of regression and other correlation-based models provides us with important insights into the consequences that are associated with, and might even flow from the actions that we have taken in the past. These models are especially useful with regard to our efforts to identify patterns of behavior that generated disparate outcomes that we may identify as being the products of objectionable, or even illegal discrimination.

Those who doubt the causal evidence brought to bear through the retrospective analysis of correlations may on occasion be persuaded by the introduction of additional variables representing alternative causal hypotheses, or reflecting the influence of particular circumstances.

Some, however, still argue that the demonstration of causal influence outside the idealized setting of experiments with random assignment is impossible. Part of this difficulty is due to the fact that many of the factors that influence outcomes of interest, like health status, are themselves effects. Their inclusion in regression models in an effort to identify the important, and independent sources of influence is ironically self-defeating because, as of yet, there is no way to truly separate out the contribution of many factors that work together in complex ways (Oakes 2004).

Our ability to increase our confidence in the explanations of the past that these models provide by adding additional data is something that is generally not available to us when similar models are put into service as predictors of the future, or as justifications for future courses of action. Additional data from the future are simply not available for us to use in the present.

The best we can do is to engage in some informed, but still speculative simulation of possible futures. The simulations may allow us to examine how relations in our imagined future worlds change in response to the introduction or adjustment of the parameters of variables we believe to be important. In some models, the parameters are allowed to vary randomly, or in accordance with some familiar stochastic function. The fact that the various states of the future, including those reflected in variables not included in our model are simply not available to us in the present will always contribute to the problem of misspecification. Data from simulations suffer from the “problem of the second best.” While we know that these data are not as good as “the best,” we have no way of knowing just how much less accurate they are.

Especially troublesome is the fact that critical aspects of the future will have been shaped in the present and in the past by actors who have relied on their own incomplete models of the future. Although Giddens does not refer to this kind of challenge specifically, it fits well within his category of “unacknowledged conditions of action” that serves as a constraint on human agency (Giddens 1984: 5–7).

The Rhizomatic or Viral Spread of Probability and Prediction

Gerd Gigerenzer and his colleagues (Gigerenzer et al. 1989) have provided an expansive, but necessarily incomplete catalog of the great variety of ways in which probability and statistics have come to play such a powerful role in the modern world. They focus most of their attention on the overlapping spheres of activity that are defined primarily by the use of applied statistics. Their review helps us to appreciate the ways in which the “benefit of the doubt” granted to experts associated with scientific approaches to knowledge is amplified in rather dramatic ways when that expertise is expressed in mathematical terms. In part, because of the rather low level of numeracy observed in the US population, mathematical formulations have achieved special, nearly mystical social status (Hobart and Schiffman 1998), setting aside other traditional ways of stating problems, and balancing competing interests.

As they suggest, it seems like each day provides a reminder that “probability theory has become the arbiter of practical rationality, not merely its mathematical codification” (Gigerenzer et al. 1989: 255). With scientific research at its core, a sociotechnical infrastructure of discovery enables developments in the diagnosis and treatment of disease. Often, some of the same tools and techniques are used to design the targeted campaigns to market the drugs, as well as to evaluate the economic impact of extending the reach of the patent and copyright shields their developers have petitioned the government for. Facility with the use of statistical

tools has become an essential component of any claim of expertise within most professional domains.

In his analysis of the prison, and the role of disciplinary surveillance more generally, French philosopher Michel Foucault (1979) called attention to the rise of a cadre of experts to whom the responsibility for observation, evaluation, and classification was granted and then jealously guarded. Although he relied upon the architecture of an idealized prison as a metaphor for the social technology he sought to describe, Foucault was quite clear that panoptic schema would come to be applied in an ever-expanding variety of institutional settings that may only seem like prisons to those struggling under the weight of expertise (Lyon 2001).

Risk and Governmentality

The primary application of the technologies that invites our attention and concern in this book has to do with the estimation and management of risk. Concern about risk has come to replace our interest in discovering the essence of the good life with an attempt to identify and control those things that threaten to put it out of reach.

An important contribution to our thinking about risk, and its emerging centrality in our thinking about society are the contributions from Anthony Giddens (1990) and Ulrich Beck (1992). Although Beck's initial contribution has come in for some criticism, there is little doubt that his book about the "Risk Society" and the limits of expertise has helped to challenge our thinking about the role of planning and analysis (Mythen 2007).

Beck's critical assessment of the unintended consequences of technology emphasized the extent to which these new hazards exceed the capacity of contemporary systems of regulation to either predict or to minimize their impact. However, many critics have challenged Beck's assumptions regarding the expansion of risk as a "universalizing" process that somehow erases traditional divisions along lines of race and class.

For many, there are structural features of society that give racism and poverty a common underlying force. Like magnetism, this force attracts hardship and suffering in ways that reinforce the boundaries of inequality and class (Mythen 2007: 799–800). Even Beck recognized that some of the ways in which poverty and race attract hardship captures public attention when the mass media can no longer avoid displaying the evidence. Such was the case with the masses of poor African Americans who suffered in the wake of Katrina's floodwaters (Beck 2006: 339). Yet, from Beck's perspective on modernization, it was actually the welfare state that produced the "individualization of social inequality" that helped to dissolve the meaning of large group categories like race and class (Atkinson 2007).

In the same way that the individualization of personality proceeds through a kind of "do-it-yourself" construction of identity, the risk regimes that Beck

describes also invite individuals to choose their own levels of exposure to a variety of risks. Many of these exposures to risk are associated with a continually varying array of choices regarding employment. Unfortunately, Beck still seems to ignore the fact that options within labor markets are still rather firmly structured along corridors defined by race, ethnicity, and class (Mythen 2005).

Another aspect of the risk society that owes a debt to Beck's analysis is the tendency of governments, especially in those in the US and the UK, to explicitly encourage the privatization of risk through a systematic reduction in financial and administrative support for public agencies and programs. Many of these "downsized" programs previously provided something of a social safety net to protect citizens in times of crisis (Calhoun 2006). Thus, in the risk society that Beck only partially understood, everyone may share the risks, but only the truly disadvantaged are likely to feel the pain.

The rise of experts in risk analysis who serve as resources for the agents of governmentality has been linked with the emergence of neoliberalism as a dominant ideological framework (O'Malley 2000). The regulatory focus of the management of risk was in one sense, a counter-weight to the movements toward deregulation, and a narrowing of the range of government activities that were not opened up to private enterprise. The result was a form of risk-based regulation that relied upon techniques such as cost/benefit analysis.

The importance of actuaries, and actuarial thinking inside and outside of government is just beginning to be understood. The actuarial assumption is based on the belief that patterns observed in the past provide useful information about activities and events that may take place in the future. This is the assumption behind the use of statistical analysis that reveals the associations between measured attributes of actors, and features of the environments in order to make predictions about how similarly situated actors will behave in the future. The assumption is that there is sufficient information about the future to be inferred from previously "observed" relationships for analysts to make statements about what the pattern of relationships in the future are likely to be. What is routinely ignored is the fact that predictions are made regarding averages for aggregates, rather than about any particular individual case.

The criminal justice system (CJS) is just one of the institutional sites in which thinking about risk and the ways it could be managed has been transformed by actuarial thinking. Whereas prior thinking focused on "deterrence, correction and the social and psychological causes of crime," an actuarially based approach to crime control increasingly focuses on the reduction of "situational opportunity" (O'Malley and Hutchinson 2007: 374–5).

Another important arbiter of opportunity that has accepted actuarialism into its bag of analytical and predictive tools is the community of economists who are called upon, or "volunteer" their insights on matters of rational choice. Although the history of their authoritative rise is far from complete, the elevation of economic analysis to a position of dominance in law and social policy has been examined by a number of scholars (McCloskey 1990, Reder 1999).

Many of us find it relatively easy to accept the expertise of economists as it relates to their presentations of historical and “institutional facts” about the economy that they routinely provide like accountants merely reviewing the books. A bit more caution seems in order when economists are relied upon as guides to economic and social policy. Their recommendations have been subject to especially harsh criticism when they have ventured into markets that are far removed from the idealized markets assumed into being within received theory (Reder 1999: 279–85). A good example can be found in the some of the contributions that economists have made to policy debates within the Federal Communications Commission (Baker 2002).

It is when economists stray even further beyond the borders of what we might ordinarily accept as the domain of economic problems that doubt and concern begins to trouble our sleep. It is when they being to offer policy guidance on a broad range of policy matters based on problematic assumptions about efficiency and rationality that those concerns become a constant distraction (Sunstein 1997). A good number of observers have been especially critical about the “disciplinary imperialism” of the brand of economic analysis associated with The Chicago School, and its preference for the rule of the market, even where no well functioning markets could possibly exist (Aune 2001). But public acceptance of this kind of social expertise seems likely to grow in the wake of an outpouring of popular acclaim for *Freakonomics*, and its data driven approach to examining the “hidden side of everything” (Leavitt and Dubner 2005).

In some areas, the rise of statistical thinking has not always been smooth. Among the barriers limiting the widespread acceptance of statistical evidence and expertise in the nation’s courts has been the lack of conventional standards for assessing claims of statistical confidence, and causal influence. The level of uncertainty and disagreement among the experts offering testimony before the courts has only served to confuse the finders of fact, as well as those who would advise them about how they should proceed (DeGroot et al. 1986).

Evaluating Social Technologies

When we think about the development and diffusion of new technologies into routine use by business and government, the primary focus of our attention is usually on the efficiency and effectiveness with which these devices and techniques can contribute to the bottom line, or contribute to the expanded production of some social good. Far too often we tend to ignore the ways in which these benefits are being distributed. And, because our traditional focus on maximization tends to emphasize benefits, the distribution of burdens is hardly considered at all.

We tend to think about technologies in terms of the benefits they provide; how they extend our reach or control beyond the limits of the human body. We spend far less time thinking about the negative, and largely unintended consequences that flow from the use of these resources (Tenner 1996). Similarly, we tend to give

the benefit of the doubt to proposed technologies on the basis of the value we place on the goals they are meant to serve.

If pressed, as a society we might rank order social or institutional users and uses of technology along some continuum of desirability. The police and their efforts to control crime would probably rank high on such a list, and we might grant them the benefit of the doubt about the technologies that they call into use. The techniques that we put to use when we are engaged in the pursuit of less highly valued goals may find less support, even though they may be indistinguishable from those used for more apparently righteous ends.

Of course, we understand that these technologies do not drop from the sky fully formed, and ready for use. They are developed over time, and they are developed in response to the demands of institutions and actors within segments of the economy that have sufficient resources to attract the attention and dedication of developers.

Nearly all of these tools and techniques are discriminatory technologies. They are used to identify, classify, and evaluate different entities or objects. These objects are evaluated primarily in terms of their expected benefits or costs. Because their evaluation is increasingly being made in the context of uncertainty or risk, these systems are increasingly oriented toward providing guidance for actions to be taken in the future. These actions almost always involve making choices among similarly situated options, and the options that should matter the most are the options that involve choices among people, or the places where people will make their lives.

Just as we noted with regard to the use of statistics to explain the past, a continuing challenge in the use of statistics in support of actions to be made in the future is the selection of the variables and data that will be used within the models. The variables that eventually come to be used in these models are not selected entirely on the basis of some unbiased, objective process that is guaranteed by some official stamp of approval, or even by a process of blind peer review.

There is no recipe or standard routine that governs the selection of variables for inclusion in classificatory schemes. There is no rigorously applied standard that establishes a cut-off point for the acceptance of models beyond the often-misconstrued standard of “statistically significant.” The fact that significance can often be achieved with a larger sample, even though the amount of variance explained remains fractional should not be ignored. We should not be surprised how often it actually is.

There is always a confluence of actors and interests that influence the selection of variables for use in analytical systems. Other factors, including the influence of dominant theories about the nature of the underlying process may also govern the selection of candidate variables. Candidate variables may also emerge from data mining analyses of records systems that become available, perhaps through the merger of corporations that were in previously independent fields of activity.

We may understand, but we soon forget the fact that the data used in these models are actually the products of the theoretical concepts that define a

methodology. Data in use are just a subset of countless possible representations of what might be the truth, reality or essence of some entity or phenomenon. All data are abstractions, and more often than we really appreciate, they can only be partial approximations of the reality they are said to represent (Bogle 2009).

While concerns about the selection of variables, and the data that are chosen to represent them ultimately affect the accuracy, precision and reliability of the models that use them, these concerns also relate to the ways in which systematic biases may be introduced. Some of these biases that have been introduced into both explanatory and predictive models reflect the influence of self-interest (Ioannidis 2005b).

Applications that Evoke Concern

This book is concerned about the consequences that flow from the use of probability and statistics in support of discrimination, rational or otherwise. It argues that the use of discriminatory technologies contributes to the social, economic and political disparities that continue to assign African Americans and other poor people of color to the fringes of society, and condemns many of them to a life of extreme relative deprivation.

Cumulative disadvantage refers to the ways in which historical disadvantages cumulate over time, and across categories of experience (Hannon 2003). The notion of cumulative disadvantage helps to explain how a racial effect can be produced within a society that may have in fact experienced a decline in the level of animus or negative racial intent as the motivation behind critical choices that have been made.

For example, while racial animus may have been at the base of discriminatory lending decisions during the initial growth in home ownership in the US, economic rationality is sufficient to explain many of the financial barriers and burdens that African Americans still face in the housing market. Because of the ways in which racial bias shaped estimates of the present and future value of real estate in neighborhoods with a substantial black presence, contemporary estimates of loan value and risk reflect and reinforce those views, even though the incorporation of racial measures into those calculations is barred by law (Squires 2003).

Each chapter in this book will examine the ways in which statistical techniques enable the generation of invidious distinctions between people and places in ways that contribute to the continually declining status of an expanding segment of the truly disadvantaged.

For many of the reasons that have already been noted, including the certainty of error, this book is intended to raise the level of concern, perhaps even to the point of alarm, about the consequences that flow from the routine use of analytical software in support of rational or statistical discrimination. The applications that have been selected for emphasis in this book have been chosen because their

contributions to the cumulative disadvantages that are helping to tear the social fabric in the US and around the world are too important to be ignored.

Market Segmentation

While risk classification assists the providers of goods and services with information about which consumers to avoid, risk avoidance is not the dominant framework through which business users approach the analysis of their markets. Unlike risk classification, market segmentation techniques are designed to sort and rank order present or potential customers in terms of their projected lifetime contributions to the bottom line.

Sophisticated modeling supports the development of targeted communications that are designed to activate the desired responses in narrower and narrower segments of the market. While segmentation after the fact will allow firms to provide specially tailored offers to customers that seek them out, predators are more likely to seek out their victims and press their offers on them, often under the cloud of dire warnings about rapidly declining opportunity.

Market segmentation is combined with target marketing communications to match the presentation of offers with the highest probability of success. Increasingly, sophisticated programs that identify market segments are also relied upon for recommendations about the prices to be charged to members of each segment. Although consumers tend to resent this kind of price discrimination when they become aware of it, most consumers' decisions are made on the basis of incomplete information about the nature of price and quality variation in the market (Edwards 2006).

Increasingly market segmentation is being organized spatially. With the assistance of geodemographic cluster analysis, marketers and investors act in ways that shape the quality of the lived environment (Parker et al. 2007).

Market segmentation and targeting have also been brought into active use within the public sphere. Political parties and their consultants have used massive databases to support "micro-targeting" of potential swing voters in states where their analysis suggests that contests have a chance to be won. Predictive models in common use allows electoral campaigns to estimate the cost for swing voter shifts produced by different communication strategies (Hillygus and Shields 2008).

Data mining is a generalized approach to the discovery of patterns in information. Often, the source information has been captured from transaction-generated data. Data mining has emerged as an important resource for the segmentation of markets into risk and value based categories (Danna and Gandy 2002). Because market segmentation involves categorical discrimination, and because the groups or categories need not be the equivalent of politically protected groups, traditional restraints on discrimination may not apply (Chung and Grimes 2005). For example, while "redlining" maybe forbidden within markets for real estate, or for associated services, such as mortgage lending, "weblining" on the basis of virtual community classification faces no such legal constraints.

Legal scholar Daniel Solove (2008) raises an important concern about the difficulty people face in defending their rights against strategies based on data mining. While they may successfully challenge claims made about their culpability by demonstrating that it was impossible for them to have done what was claimed, it is far more difficult for them to demonstrate that they would not, or will not do what some predictive model suggests.

The same technologies that are used in support of legitimate economic and social goals are also being put to use by those motivated by criminal intent. Predictive models are used to identify likely victims in the same ways in which they might be used to identify potential customers or unacceptable risks. In markets for consumer credit, predatory lenders engage in a form of “reverse redlining” where homeowners who were at the greatest risk of default were being targeted (Howell 2006, Willis 2006). Consumer fraud captures a substantial part of the nation’s economic resources, and the government’s ability to control it lags far behind the development and spread of the new strategies and tactics being used.

Assessments of Insurability

The provision of insurance products and services is arguably the most widespread and influential application of probability and statistics today. Unfortunately we are just coming to realize that the institution of insurance has been engaged in a pervasive form of “moral regulation” that extends far beyond the boundaries suggested by estimates of annual premium income (Baker 2000).

Debates about the importance of avoiding the “moral hazard” the economy would face if all Americans were not required to purchase health insurance represent a mere fraction of the ways in which the moral force of insurance has been felt. Among the most important ways in which the technology and logic of insurance extended its reach is through the industry’s success in promoting its own interpretation of “actuarial fairness” (Thomas 2007).

The metaphoric construction favored by private insurance carriers has helped to justify an analytical arms race within the industry. Competitors raced to develop the most sophisticated techniques for risk classification so that the rates people are charged will more accurately match the lifetime value of the claims they would submit, unless, of course, they were denied coverage, or priced out of the market (Crossley 2005).

Policy debates about restrictions on the ability of insurers to gather and use personal information including genetics to set rates often fail to include realistic estimates of the ability of consumers to use their so-called informational advantage to the detriment of insurers as a group. While it is easy to imagine this kind of abuse by consumers in the market for life insurance, it makes little sense in the market for health insurance (Chandler 2002), or in the other emerging markets in which uncertainty rules through fear.

Policing and Crime Control

In 2008, a truly an embarrassing status was achieved by the United States. The rate of imprisonment for adults exceeded one in 100 for the first time in American history. No other nation matched the rate, or the total number of persons behind bars. This historic rate actually increases to one in 30 for young men. For African Americans, the rate rises to one out of nine (Warren 2008).

The differential risk of being imprisoned reflects the influence of an array of factors, including the widespread use of statistically informed predictions of dangerousness and likelihood of offending (Taxman et al. 2005). Predictive modeling that informed decisions about which prisoners should be eligible for release became more important as the rising costs of the prison system threatened the economic stability of many state governments.

Other statistical techniques merely add to the numbers of persons at risk. Data mining and other means of computer enhanced surveillance lead to an expansion in the number of potential targets. Because of declining costs of capture, storage, and processing of transaction generated information (TGI) lower standards of potential harm become normalized, more and more people are placed at risk of arrest for a broadened array of prosecutable offenses. Administrative guidelines that once established cost “triggers” at \$10,000 for heightened scrutiny of financial transactions at banks could be lowered to amounts of \$500, or even \$100 in the foreseeable future. When the application of these data mining and surveillance techniques moves the goal of policing beyond investigation and toward prevention, perhaps through stings or pre-emptive strikes, our basis for concern about systematic bias is almost certain to expand (Balkin and Levinson 2006).

The police have also revealed considerable interest in the uses to be derived from advances in biometrics. Biometrics refers to the rapidly development of a science of identification that relies upon the analysis of an individual’s physiological or behavioral characteristics. Biometrics are attractive as means of identification because unlike familiar tokens that can be misplaced, stolen, or modified, these identifiers cannot be readily shared or left behind.

Biometric systems that capture images of the face, eyes, hands and fingers as well as signatures and samples of spoken words may be combined with other tokens of identification in order to enable authorization for some critical, or increasingly routine transaction. Substantial investments in the development of biometric technologies are driven by the presumed need to make identification and authorization automatic, or at the very least, accomplished reliably with minimal human effort.

Fingerprints, DNA, and surveillance cameras all rely to some degree on probabilistic assessments of the probability that the person “identified” by the system is in fact the person whose “data” have been left behind (Gates 2005). The problem of identification goes beyond the association of a unique individual with a reliable indicator; it is also bound up within the challenge of correctly placing

that individual within a great variety of categories or classes whose members are ranked and rated against some standard or criterion measure (Joh 2006, Phillips 2004).

Like estimates of probability, biometric systems usually provide some kind of index of similarity, or the degree of match between the reference image and the image presented by the person being identified. Precision, or confidence increases with the number of points of comparison. As with other tests based on statistics, samples, and distributions, errors of over- or under-identification are assessed and weighted differently on the basis of the nature and distribution of costs associated with each type of error.

Genetic profiling in the area of criminal law has been a particularly active area of interest (Rothenberg and Wang 2005). There are important distinctions to be drawn between the uses of DNA for identification, and its uses for behavioral classification and predictions of dangerousness (Rothstein and Talbott 2006).

The Role of the Media

Because of the role that the mass media play in shaping the public's understanding and concern about the ways in which cumulative disadvantage threatens the social fabric and potential of the nation, this book also explores the ways in which probability and statistics have weakened the ability of the media to serve the public interest in the ways we have assumed it should.

Stories about social problems, including investigative series about racial disparity have become regular features in news and public affairs. The electronic media routinely make use of statistical data and analysis to convey impressions of disparity to their audiences. Newspapers rely on colorful graphics in an effort to facilitate public understanding of trends, patterns, and relationships. Even multi-colored maps have become familiar devices for revealing the ways in which quality of life indicators might vary across markers of time, space and community.

Unfortunately, the media organizations that are now providing their audiences with more compelling displays of what they assume is a statistical reality, pay far too little attention to the origins of the data or the representations that they feature so prominently. Most of the political actors, and members of the politically engaged electorate need to be reminded that most of the statistics we encounter in the media are there because the people providing them are hoping to be able to persuade the public about something (Best 2001, Gandy 1982).

Although much of the misrepresentation of social reality that is distributed by the media is the product of well-planned strategic efforts, there are other, more innocent ways in which bias and distortions come to be distributed as fact by the media. Often, news stories that make use of summary statistics to make direct comparisons between the winners and losers in the game of life tend to focus on the losers, rather than the winners (Gandy and Li 2005). Ironically, by focusing on the victims, rather than on those who are advantaged by race, gender, or social

position, journalists may actually have contributed to the reproduction of the conditions that they set out to change through their efforts.

The Design of the Book

The following chapters will examine in greater detail the ways in which the use of probability and statistics as an aid to decision-making by institutions of business and government have contributed to the cumulative disadvantage of the poor, especially African Americans.

Chapter 2 explores the history and development of probability and statistics as applied social technology. It includes a critical assessment of the limits of predictive models, based in part on the nature of data, as well as on the basis of assumptions about the reliability of predictions. It includes an extended discussion of the factors that limit the choice of variables used in commercially available, as well as proprietary analytical models. It also explores some of the moral and ethical conflicts that are raised by consideration of the role played by chance in the choices that are made.

Chapter 3 explores the nature of genetic differences. It underscores the probabilistic nature of genetic endowments, and the relationship between genetic and environmental influences on health, behavior, and personality. It emphasizes the role that chance plays in establishing the relationship between identification, classification, and outcomes in the presence of genetic markers. It focuses in particular on the active pursuit of genetic markers as predictors of advantage and disadvantage in life. It examines the rising level of concern about the return of eugenics and its implications for social policy formation.

Chapter 4 explores the meaning and importance of rational discrimination. It begins with an examination of the logic behind the use of statistical classification, evaluation, and prediction, and the extent to which its use is defended as being instrumentally rational. It also explores the counter arguments that propose restrictions on the use of some variables in discriminatory techniques because of the ways in which they lead to overuse and cumulative disadvantage.

Chapter 5 examines numerous examples of the ways in which statistical analysis and predictive modeling generates disparities in education, housing and access to health care. It demonstrates the ways in which marketplace discrimination generates social outcomes similar to the products of institutional racism, even though they need not be based on any kind of racial animus. An examination of so-called neighborhood effects is used to illustrate the ways in which disparity cumulates into disadvantage.

Chapter 6 examines ways in which the actuarial assumptions regarding the predictive utility of past measures, and group estimates, are used to make and justify discriminatory choices within financial markets. It examines the recent financial crisis and its origins in the subprime mortgage market to illustrate the limitations of risk rating techniques. It also includes examples of the ways in

which the classification of individuals, groups, and communities influences not only investment, but also the targeting of predatory efforts against members of high-risk groups.

Chapter 7 explores problems within the criminal justice system that flow from the use of profiles, predictions of dangerousness, and a variety of techniques to allocate scarce resources more efficiently. Differences in arrest, conviction and sentencing are discussed in terms of the role that statistical analysis plays. An assessment of the cumulative impact of predictive modeling in policing emphasizes the dangerousness of encounters with police, and the way that police records restrict future opportunities for participation in society.

Chapter 8 examines the role of risk assessment in the formation and evaluation of public policy. It includes a discussion of the role of expert analysis and prediction in the identification of problems, the evaluation of policy options, and assessments of their success. It also discusses the ways in which options are framed within legislative and judicial environments.

Chapter 9 examines the role of the media in shaping public understanding as well as the policy system's response to a range of societal problems. It includes a discussion of the influence of sources and their use of preferred frames as strategic resources in policy debates. The nature of influence is also explored in relation to the delivery of information subsidies and the sponsorship of particular policy frames. Case studies, including racial profiling, and the framing of inequality more generally are presented as illustrations of the social construction of risk. An analysis of racially comparative risk is used to reveal how unintended consequences may flow from the media's traditional focus on victims, rather than beneficiaries of public policies

The final chapter explores the need for a social movement to oppose expanded use of statistical techniques for the identification, classification and evaluation of individuals in ways that contribute further to their comparative disadvantage. It identifies a path that might be taken to limit, and perhaps to even reverse the trend toward actuarialism and statistical discrimination.